Understanding Diffusion Models: A Unified Perspective

Year: 2022 | Citations: 447 | Authors: Calvin Luo

Abstract

Diffusion models have shown incredible capabilities as generative models; indeed, they power the current state-of-the-art models on text-conditioned image generation such as Imagen and DALL-E 2. In this work we review, demystify, and unify the understanding of diffusion models across both variational and score-based perspectives. We first derive Variational Diffusion Models (VDM) as a special case of a Markovian Hierarchical Variational Autoencoder, where three key assumptions enable tractable computation and scalable optimization of the ELBO. We then prove that optimizing a VDM boils down to learning a neural network to predict one of three potential objectives: the original source input from any arbitrary noisification of it, the original source noise from any arbitrarily noisified input, or the score function of a noisified input at any arbitrary noise level. We then dive deeper into what it means to learn the score function, and connect the variational perspective of a diffusion model explicitly with the Score-based Generative Modeling perspective through Tweedie's Formula. Lastly, we cover how to learn a conditional distribution using diffusion models via guidance.